SKEW-NORMAL/INDEPENDENT DISTRIBUTIONS, WITH APPLICATIONS

V. H. Lachos and F. V. Labra

Departamento de Estatistica, Universidade Estadual de Campinas
Caixa Postal 66281 - CEP 05315 970
Campinas - SP - Brazil

Abstract

Normal/independent distributions are often used as a challenging family for statistical procedures of symmetrical data. In this article, we have defined a skewed version of these distributions in the multivariate setting and we have derived several of its properties. The main virtue of the members of this family of distributions is that they are easy to simulate from and they also lend themselves to the Monte Carlo EM algorithm for maximum likelihood estimation. For multivariate skewed responses, the EM-type algorithm has been discussed with emphasis on the skew-t, on the skew-slash, and on the skew-contaminated normal distributions. Results obtained from simulated and real data sets are reported illustrating the usefulness of the proposed methodology.

Key Words: *MC-EM algorithm; normal/independent distributions; skewness.*

1 Introduction

The normal/independent distributions (Lange and Sinsheimer, 1993) provide a group of thick-tailed distributions that are often used for robust inference of symmetrical data. The theory and applications generate a great number of data that are skewed or heavy-tailed, for instance, the data of family income. Thus, we need appropriate distribution to fit and simulate these skewed or heavy-tailed data. Candidate distributions at our disposal for fitting and simulating these data are not very abundant in the literature. In this article, we propose a new family of distributions that combine skewness with heavy tails. Moreover, this distribution is attractive because it has a stochastic representation that allows easy implementation of the EM-algorithm and it also facilitates the study of many of its properties . Our proposal generalized recent results found in Gupta (2003) and Wang and Genton (2006).

A simpler departure which defines the univariate skew-normal distribution has been proposed by Azzalini (1985). An extension to the multivariate setting was proposed by Arellano–Valle et al. (2005) (see also Azzalini and Dalla- Valle, 1996), defining the following probability density function (pdf)

$$f(\mathbf{y}) = 2\phi_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))\,, \quad \mathbf{y}\in\mathbb{R}^p. \tag{1}$$

where $\phi_p(.|\boldsymbol{\mu},\boldsymbol{\Sigma})$ stands for the pdf of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$, $\Phi_1(.)$ represents the cumulative distribution function (cdf) of the standard normal distribution, and $\boldsymbol{\Sigma}^{-1/2}$ satisfies $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} =$

$\mathbf{\Sigma}^{-1}$. When $\boldsymbol{\lambda} = \mathbf{0}$, the skew normal distribution reduces to the normal distribution ($\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$). A p-dimensional random vector $\mathbf{Y}$ with pdf as in (1), will be denoted by $\mathrm{SN}_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda})$. Its marginal stochastic representation, which can be used to derive several of its properties, is given by

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{\Sigma}^{1/2}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1), \quad \text{with} \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}, \tag{2}$$

where $|T_0|$ denotes the absolute value of T_0, $T_0 \sim \mathrm{N}_1(0, 1)$ and $\mathbf{T}_1 \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$ are independent, and " $\stackrel{d}{=}$ " means "distributed as". From (2) it follows that the expectation and variance of $\mathbf{Y}$ are given, respectively, by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\mathbf{\Sigma}^{1/2}\boldsymbol{\delta}, \tag{3}$$

$$Var[\mathbf{Y}] = \mathbf{\Sigma}^{1/2}(\mathbf{I}_p - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^\top)\mathbf{\Sigma}^{1/2}. \tag{4}$$

Following the same ideas of Azzalini (1985), it is now natural to construct univariate and multivariate distributions that combine skewness with heavy tails. For instance, one can define skew-t distributions (Sahu et al., 2003; Gupta, 2003), skew-Cauchy distributions (Arnold and Beaver, 2000), skew-slash distributions (Wang and Genton, 2006), skew-slash-t distributions (Tan and Peng, 2006), skew-elliptical distributions (Azzalini and Capitanio, 1999; Branco and Dey, 2001; Sahu et a., 2003; Genton and Loperfido, 2005). In this article, we define a new family of asymmetric distributions and we study its properties and applications. This new family contains the multivariate skew-normal distribution defined by Arellano-Valle et al. (2005), the multivariate skew-slash distribution defined by Wang and Genton (2006), the multivariate skew-t distribution defined by Gupta (2003), and all the distributions studied by Lange and Sinsheimer (1993) in a symmetric context.

The paper is organized as follows. In Section 2, for the sake of completeness, we give a brief sketch of normal/independent distributions (NI). In Section 3, the skew-normal normal/independent distributions (SNI) are defined by extending the NI models. Properties like moments, linear transformation and stochastic representation of the proposed distributions are also discussed. In Section 4, an Monte Carlo EM-type (MC-EM) algorithm which presents advantages over the direct maximization approach is presented, especially in terms of robustness with respect to starting values. Section 5 reports applications to simulated and real data sets, indicating the usefulness of the proposed methodology. Concluding remarks are given in Section 6.

2 Normal/independent distributions

The symmetrical family of NI distributions has attracted attention in the last few years, particularly due to the fact that they include distributions such as the Student-t, the slash, the power exponential, the contaminated normal. All of these distributions have heavier tails than the normal ones. We say that a p-dimensional vector

$\mathbf{Y}$ has an NI distribution (see for instance, Lange and Sinsheimer, 1993) with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$ and a positive definite scale matrix $\boldsymbol{\Sigma}$ if its density function assumes the form

$$f(\mathbf{y}) = \int_0^\infty \phi_p(y|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})dH(u), \tag{5}$$

where $H(u;\boldsymbol{\nu})$ is a cdf of a unidimensional positive random variable U indexed by the parameter vector $\boldsymbol{\nu}$. For a random vector with a pdf as in (5), we shall use the notation $\mathbf{Y} \sim \mathrm{NI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$. Now, when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$, we shall use the notation $\mathbf{Y} \sim \mathrm{NI}_p(H)$. Its stochastic representation is given by

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \tag{6}$$

where $\mathbf{Z} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ and U is a positive random variable with cdf H independent of $\mathbf{Z}$. Examples of NI distributions are described subsequently (see Lange and Sinsheimer, 1993). For this family, the distributional properties of the Mahalanobis distance

$$d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

are described, because they are extremely useful in testing the goodness of fit and detecting outliers.

2.1 Examples of NI distributions

- *The Student-t distribution with $\nu > 0$ degrees of freedom, $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$.* The use of the t-distribution as an alternative to the normal distribution, has frequently been suggested in the literature, for example, Little (1988) and Lange et al. (1989) use the Student-t distribution for robust modeling. $\mathbf{Y}$ has a density given by

$$f(\mathbf{y}) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} (1 + \frac{d}{\nu})^{-(\frac{p+\nu}{2})}. \tag{7}$$

  In this case, we have that $U \sim Gamma(\nu/2, \nu/2)$, where $H(u;\nu)$ has density

$$h(u;\nu) = \frac{(\nu/2)^{\nu/2} u^{\nu/2-1}}{\Gamma(\nu/2)} \exp{(-\frac{1}{2}\nu u)}, \tag{8}$$

  with finite reciprocal moments $E[U^{-m}] = \dfrac{(\nu/2)^m \Gamma(\nu/2 - m)}{\Gamma(\nu/2)}$, for $m < \nu/2$. From Lange and Sinsheimer (1993), it also follows that

$$d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sim pF(p, \nu).$$

- *The slash distribution, $\mathbf{Y} \sim SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, with a shape parameter $\nu > 0$.* This distribution presents heavier tails than those of the normal distribution and it includes the normal case when $\nu \uparrow \infty$. Its pdf is given by

$$f(\mathbf{y}) = \nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})du. \tag{9}$$

Here we have that $H(u;\nu)$ has density

$$h(u;\nu) = \nu u^{\nu-1}\mathbb{I}_{(0,1)}, \tag{10}$$

with reciprocal moments $E[U^{-m}] = \dfrac{\nu}{\nu - m}$, for $m < \nu$, and the Mahalanobis distance has cdf

$$Pr(d \le r) = Pr(\chi^2_p \le r) - \frac{2^\nu\Gamma(p/2+\nu)}{r^\nu\Gamma(p/2)} Pr(\chi^2_{p+2\nu} \le r).$$

- *The contaminated normal distribution,* $\mathbf{Y} \sim CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \gamma)$, $0 \le \nu \le 1$, $0 < \gamma \le 1$ (Little, 1988). This distribution may also be applied for modeling symmetric data with outlying observations. The parameter ν represents the percentage of outliers, while γ may be interpreted as a scale factor. Its pdf is given by

$$f(\mathbf{y}) = \nu\phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{\gamma}) + (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{11}$$

In this case the probability function $H(u;\boldsymbol{\nu})$ is given by

$$h(u;\boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \;\; \boldsymbol{\nu} = (\nu,\gamma)^\top, \tag{12}$$

where the notation $\mathbb{I}_{(A)}$ is the indicator function of the set A. Clearly, $E[U^{-m}] = \nu/\gamma^m + 1 - \nu$, and

$$Pr(d \le r) = \nu Pr(\chi^2_p \le \gamma r) + (1-\nu)Pr(\chi^2_p \le r).$$

The power-exponential distribution is of the type NI. However, the scale distribution $H(u;\boldsymbol{\nu})$ is not computationally attractive and it will not be dealt with in this work.

3 Skew-normal/independent distributions

In this section, we define the multivariate SNI distributions and study some of its properties. We have also shown that these distributions are invariants under linear transformations.

Definition 1. *A p-dimensional random vector $\mathbf{Y}$ follows an SNI distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, scale matrix $\boldsymbol{\Sigma}$ (an $p \times p$ positive definite matrix) and skewness parameter $\boldsymbol{\lambda} \in \mathbb{R}^p$, if its pdf is given by*

$$\begin{aligned} f(\mathbf{y}) &= 2\int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1(u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))dH(u) \\ &= 2\int_0^\infty \frac{u^{p/2}}{(2\pi)^{p/2}}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{u}{2}d_\lambda}\Phi_1(u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))dH(u), \end{aligned} \tag{13}$$

where U is a positive random variable with cdf $H(u;\boldsymbol{\nu})$. For a random vector with pdf as in (13), we use the notion $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. If $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$ we refer to it as a standard SNI distribution and we denote it by $SNI_p(\boldsymbol{\lambda}; H)$.

Clearly, from (13), when $\boldsymbol{\lambda} = \mathbf{0}$ we get the corresponding NI distribution defined in (5). For a random vector with pdf as in (13), we write the Mahalanobis distance as

$$d_\lambda = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}).$$

In definition 1, note that the cdf $H(u; \boldsymbol{\nu})$ is indexed by the parameter vector $\boldsymbol{\nu}$. Thus, if we suppose that $\boldsymbol{\nu}_\infty$ is such that $\boldsymbol{\nu} \uparrow \boldsymbol{\nu}_\infty$, and $H(u; \boldsymbol{\nu})$ converges weakly to the distribution function $H_\infty(u) = H(u; \boldsymbol{\nu}_\infty)$ of the unit point mass at 1, then the density function in (13) converges to the density function of a random vector having a skew-normal distribution. The proof of this result is similar to that of Lange and Sinsheimer (1993) for the NI case.

For an SNI random vector, the stochastic representation given below can be used to quickly simulate pseudo realizations of $\mathbf{Y}$ and also to study many of its properties.

Proposition 1. *Let $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. Then*

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \tag{14}$$

where $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ and U is a positive random variable with cdf H independent of $\boldsymbol{Z}$.

Proof. The proof follows from the fact that $\mathbf{Y}|U = u \sim \mathrm{SN}_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. □

Notice that the stochastic representation given in (6) for the NI case is a special case of (14) when $\boldsymbol{\lambda} = \mathbf{0}$. Hence, we have extended the family of NI distributions for the skewed case. Besides, from (2) it follows that (14) can be written as

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \frac{1}{U^{1/2}} \boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|X_0| + (\mathbf{I}_n - \boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{X}_1\}, \tag{15}$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$, and U, $X_0 \sim N_1(0,1)$ and $\mathbf{X}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$ are all independent. The marginal stochastic representation given in (15) is very important since it allows to implement the EM-algorithm for a wide variety of linear models similar to those of Lachos et al. (2007).

Remark 1. *If $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, then from (15) it follows that*

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|W_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{W}_1\}, \tag{16}$$

where $W_0 \sim NI_1(H)$ and $\mathbf{W}_1 \sim NI_p(H)$. In other words, if we know mechanisms for to generate of a standards NI distribution, then also we know a mechanism to generate of the related SNI distribution.

In the next proposition, we derive a general expression for the moment generating function (mgf) of a SNI random vector.

Proposition 2. *Let $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. Then*

$$M_{\mathbf{y}}(\mathbf{s}) = E[e^{\mathbf{s}^\top\mathbf{Y}}] = \int_0^\infty 2e^{\mathbf{s}^\top\boldsymbol{\mu} + \frac{1}{2}u^{-1}\mathbf{s}^\top\boldsymbol{\Sigma}\mathbf{s}}\Phi_1(u^{-1/2}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{s})dH(u), \quad \mathbf{s} \in \mathbb{R}^p \tag{17}$$

Proof. From Proposition 1, we have that $\mathbf{Y}|U = u \sim SN_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Now, from the known properties of conditional expectation, it follows that $M_{\mathbf{y}}(\mathbf{s}) = E_U[E[e^{\mathbf{s}^\top\mathbf{Y}}|U]]$ and the proof concludes of the fact that U is a positive random variable with cdf H and since, if $\mathbf{Z} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ then $M_{\mathbf{z}}(\mathbf{s}) = 2e^{\mathbf{s}^\top\boldsymbol{\mu}+\frac{1}{2}\mathbf{s}^\top\boldsymbol{\Sigma}\mathbf{s}}\Phi_1(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2})$. □

In the following preposition we derive the mean vector and the covariance matrix of a random vector SNI, the proof follows of the result in Proposition 1 and of the fact that U and $\mathbf{Z}$ are independent.

Proposition 3. *Suppose that* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Then,*

a) If $E[U^{-1/2}] < \infty$, *then* $E[\mathbf{Y}]$ *exists and*

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}E[U^{-1/2}];$$

b) If $E[U^{-1}] < \infty$, *then the second moment of* $\mathbf{Y}$ *exists and*

$$Var[\mathbf{Y}] = \boldsymbol{\Sigma}^{1/2}\left(E[U^{-1}]\mathbf{I}_p - \frac{2}{\pi}E^2[U^{-1/2}]\boldsymbol{\delta}\boldsymbol{\delta}^\top\right)\boldsymbol{\Sigma}^{1/2}.$$

Proposition 4. *If* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, *then for any even function* g, *the distribution of* $g(\mathbf{Y} - \boldsymbol{\mu})$ *does not depend on* $\boldsymbol{\lambda}$ *and has the same distribution that* $g(\mathbf{X} - \boldsymbol{\mu})$, *where* $\mathbf{X} \sim NI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$. *In a particular case, if* $\mathbf{A}$ *is a* $p \times p$ *symmetric matrix, then* $(\mathbf{Y} - \boldsymbol{\mu})^\top\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})$ *and* $(\mathbf{X} - \boldsymbol{\mu})^\top\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})$ *are identically distributed.*

Proof. The proof follows by using Proposition 2 and a similar procedure to found in Wang et al. (2004). □

As a byproduct of Proposition 4, we have the following interesting result

Corollary 1. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Then the quadratic form*

$$d_\lambda = (\mathbf{Y} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$$

has the same distribution as $d = (\mathbf{X} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$, *where* $\mathbf{X} \sim NI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$.

The result of Corollary 1 is interesting because it allows us to check models in practice (see Section 5). On the other hand, Corollary 1 together with the result found in Lange and Sinsheimer (1993, Section 2) allows us to obtain the *mth* moment of d_λ.

Corollary 2. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Then for any* $m > 0$

$$E[d_\lambda^m] = \frac{2^m\Gamma(m + p/2)}{\Gamma(p/2)}E[U^{-m}].$$

In the next proposition we shall show that an SNI random vector is invariant under linear transformations, this implies that the marginal distributions of $\mathbf{Y} \sim \mathrm{SNI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ are still SNI, with the same cdf H. This result is summarized in the following proposition:

Proposition 5. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$*. Then for any fixed vector* $\mathbf{b} \in \mathbb{R}^m$ *and matrix* $\mathbf{A} \in \mathbb{R}^{m\times p}$ *of full row rank matrix,*

$$\mathbf{V} = \mathbf{b} + \mathbf{A}\mathbf{Y} \sim SNI_p(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, \boldsymbol{\lambda}^*; H), \tag{18}$$

where $\boldsymbol{\lambda}^* = \boldsymbol{\delta}^*/(1 - \boldsymbol{\delta}^{*\top}\boldsymbol{\delta}^*)^{1/2}$*, with* $\boldsymbol{\delta}^* = (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)^{-1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$*. Moreover, if* $m = p$ *the matrix* $\mathbf{A}$ *is nonsingular, then* $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}$*.*

Proof. The proof of this result was obtained directly from Proposition 2, since $M_{\mathbf{b}+\mathbf{A}\mathbf{Y}}(\mathbf{s}) = e^{\mathbf{s}^\top\mathbf{b}}M_{\mathbf{Y}}(\mathbf{A}^\top\mathbf{s})$. When $\mathbf{A}$ is a nonsingular matrix, it is easy to see that $\boldsymbol{\delta}^* = \boldsymbol{\delta}$. □

By using (18), when $\mathbf{b} = \mathbf{0}$, we have the following additional properties of an SNI random vector.

Corollary 3. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$*. Then,*
a) $-\mathbf{Y} \sim SNI_p(-\boldsymbol{\mu}, \boldsymbol{\Sigma}, -\boldsymbol{\lambda}; H)$*;*
b) $\mathbf{a}^\top\mathbf{Y} \sim SNI_p(\mathbf{a}^\top\boldsymbol{\mu}, \mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}, \lambda^*; H)$*, for any* $\mathbf{a} \in \mathbb{R}^p$*, where* $\lambda^* = \alpha/(1-\alpha^2)^{1/2}$*, with* $\alpha = \{\mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})\}^{-1/2}\mathbf{a}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$*.*

3.1 Examples of SNI distributions

Some examples of SNI distributions, includes

- *The skew-t distribution, with* ν *degree of freedom,* $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. Considering $U \sim Gamma(\nu/2, \nu/2)$, similar procedures to found in Gupta (2003, Section 2) lead to the following density function:

$$f(\mathbf{y}) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)T_1(\frac{\sqrt{v+p}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{d+p}}|0, 1, \nu+p), \quad \mathbf{y} \in \mathbb{R}^p, \tag{19}$$

  where as usual, $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $T_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denote, respectively, the pdf and cdf of the Student-t distribution, $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, defined in (7). A particular case of the skew-t distribution is the skew-Cauchy distribution, when $\nu = 1$. Also, when $\nu \uparrow \infty$, we get the skew-normal distribution as the limiting case. See Gupta (2003) for further details. In this case, from the proposition 2 or 3, the mean and covariance matrix of $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ are given by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, \quad \nu > 1 \quad \text{and}$$

$$Var[\mathbf{Y}] = \boldsymbol{\Sigma}^{1/2}(\frac{\nu}{\nu-2}\mathbf{I}_p - \frac{\nu}{\pi}(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})})^2\boldsymbol{\delta}\boldsymbol{\delta}^\top)\boldsymbol{\Sigma}^{1/2}, \ \nu > 2.$$

- *The skew-slash distribution, with the shape parameter* $\nu > 0$*,* $SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$*.* With $h(u;\nu)$ as in (10), from the Proposition 1, can be easily seen that

$$f(\mathbf{y}) = 2\nu\int_0^1 u^{\nu-1}\phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{u})\Phi_1(u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^p, \tag{20}$$

The skew-slash distribution reduces to the skew-normal distribution when $\nu \uparrow \infty$. See Wang and Genton (2006) for further details. In this case from the propositions 2 or 3

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} \frac{2\nu}{2\nu - 1} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}, \quad \nu > 1/2, \text{ and}$$

$$Var[\mathbf{Y}] = \boldsymbol{\Sigma}^{1/2} (\frac{\nu}{\nu - 1} \mathbf{I}_p - \frac{2}{\pi} (\frac{2\nu}{2\nu - 1})^2 \boldsymbol{\delta}\boldsymbol{\delta}^\top) \boldsymbol{\Sigma}^{1/2}, \ \nu > 1.$$

- *The skew-contaminated normal distribution,* $SCN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu, \gamma)$, $0 \leq \nu \leq 1$, $0 < \gamma < 1$. Taking $h(u; \boldsymbol{\nu})$ as in (12), it follows straightforwardly that

$$\begin{aligned} f(\mathbf{y}) &= 2\{\nu \phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{\gamma}, \boldsymbol{\lambda}) \Phi_1(\gamma^{1/2} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})) \\ & \quad + (1 - \nu) \phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) \Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}))\}, \end{aligned} \tag{21}$$

in this case, the skew-contaminated normal distribution reduces to the skew-normal distribution when $\nu = 0$. Hence, the mean vector and the covariance matrix are given, respectively, by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} (\frac{\nu}{\gamma^{1/2}} + 1 - \nu) \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}, \text{ and}$$

$$Var[\mathbf{Y}] = \boldsymbol{\Sigma}^{1/2} ((\frac{\nu}{\gamma} + 1 - \nu) \mathbf{I}_p - \frac{2}{\pi} (\frac{\nu}{\gamma^{1/2}} + 1 - \nu)^2 \boldsymbol{\delta}\boldsymbol{\delta}^\top) \boldsymbol{\Sigma}^{1/2}.$$

Remark 2. *a) The stochastic representation given by equation (6) can be used to obtain the slash student distribution. Let* U_1 *have pdf as in (10),* $U_2 \sim Gamma(\nu/2, \nu/2)$, $\nu > 0$ *and* $\mathbf{X} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, *all independently distributed. Then*

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + U_1^{-1/2} U_2^{-1/2} \mathbf{X} \tag{22}$$

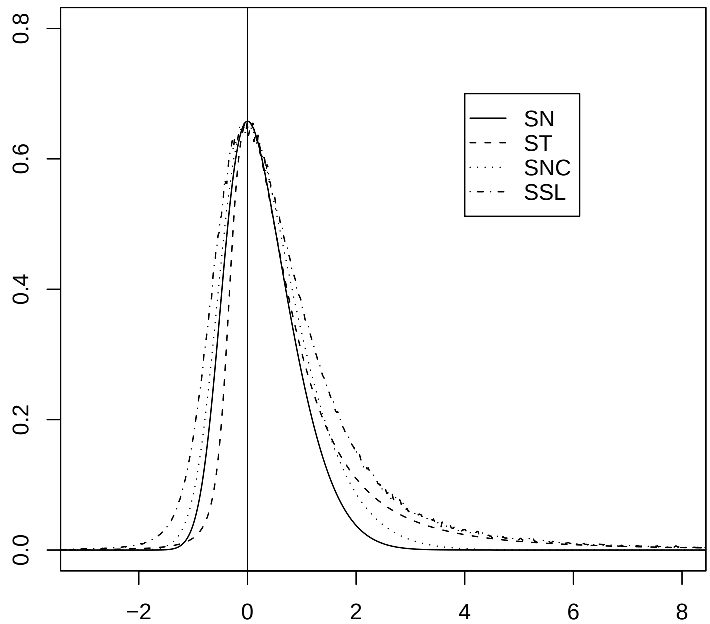
has a slash student distribution that was defined in Tang and Peng (2006). The proof follows from the fact that

$$\mathbf{T} = U_2^{-1/2} \mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu).$$

b) Now, if $\mathbf{X} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, *then* $\mathbf{Y}$ *in (22) has a skew-slash student distribution as shown by Tang and Peng (2006). Obviously, many other distributions can be constructed by choosing appropriate pdfs* ($h(.; \boldsymbol{\nu})$) *for* U_1 *and* U_2.

In Figure 1, we drew the density of the standard distribution $SN_1(3)$ together with the standard densities of the distributions $ST_1(3, 2)$, $SSL_1(3, 1)$ and $SNC_1(3, 0.5, 0.5)$. They are rescaled so that they have the same value at the origin. Note that the four densities are positively skewed, and that the skew-slash and the skew-t distributions have much heavier tails than the skew-normal distribution. Actually, the skew-slash and the skew-t distributions do not have finite means and variances. Figure 2 depicts some contours of the densities associated with the standard bivariate skew-normal distribution $SN_2(\boldsymbol{\lambda})$, the standard bivariate skew-t distribution $ST_2(\boldsymbol{\lambda}, 2)$,

Figure 1: Densities curves of the univariate skew-normal, skew-t, skew-slash and skew-contaminated normal distributions.



the standard bivariate skew-slash distribution $SSL_2(\boldsymbol{\lambda}, 1)$, and the standard bivariate skew-contaminated normal distribution $SCN_2(\boldsymbol{\lambda}, 0.5, 0.5)$, with $\boldsymbol{\lambda} = (1, 1)^\top$ for all the distributions. Note that these contours are not elliptical and they can be strongly asymmetric depending on the choice of the parameters.

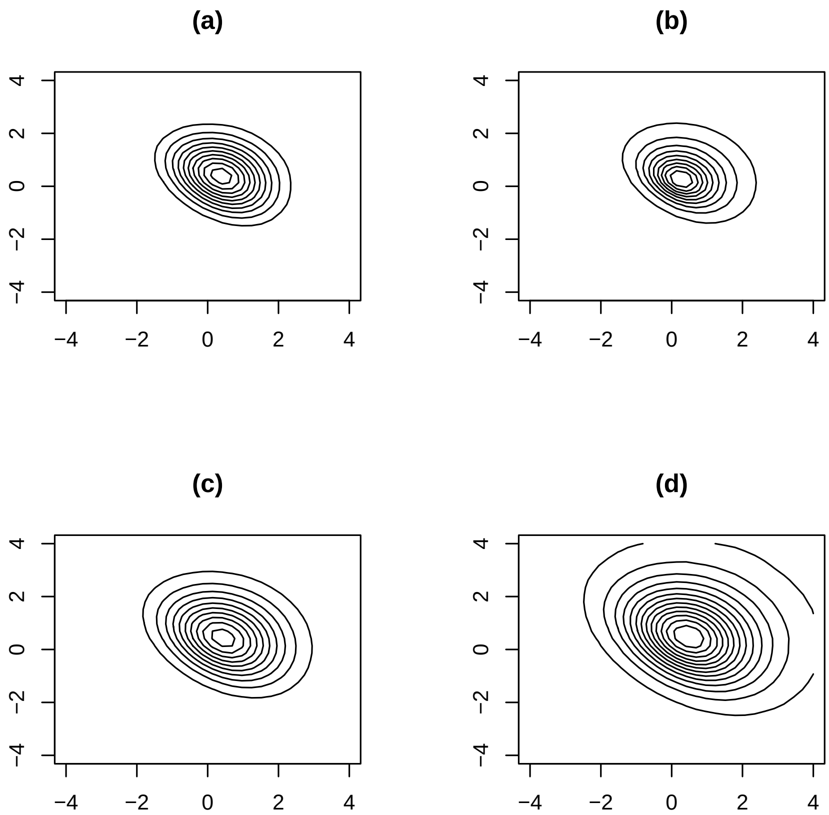
In what follows, we use the EM-algorithm in conjunction with the marginal stochastic representation given in (15) to obtain the ML estimate of the parameter vector $\boldsymbol{\theta}$. We note that it is hard to implement this approach without identifying stochastic representation. The proposed methodology does not exist even in the literature. Moreover, studies related to local influence for incomplete data (Zhu and Lee, 2004) can be easily extended from these results.

4 Maximum likelihood via the EM-algorithm

Suppose that we have observations on n independent individuals, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where $\mathbf{Y}_i \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, $i = 1, \ldots, n.$. The parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top$, where $\boldsymbol{\alpha}$ denotes a minimal set of parameters such that $\boldsymbol{\Sigma}$ is well defined (e.g. the upper triangular elements of $\boldsymbol{\Sigma}$ in the unstructured case).

In what follows, we illustrate implementation of likelihood inference via the EM-algorithm. The EM-algorithm is a popular iterative algorithm for ML estimation for models with incomplete data. More specifically, let $\mathbf{y}$ denote the observed data

Figure 2: Contour plot of some elements of the standard bivariate SNI family. (a) $SN_2(\boldsymbol{\lambda})$ (b) $ST_2(\boldsymbol{\lambda}, 2)$ (c) $SCN_2(\boldsymbol{\lambda}, 0.5, 0.5)$ (d) $SSL_2(\boldsymbol{\lambda}, 1)$, where $\boldsymbol{\lambda} = (1, 1)^\top$.



and $\mathbf{s}$ denote the missing data. The complete data $\mathbf{y}_c = (\mathbf{y}, \mathbf{s})$ is $\mathbf{y}$ augmented with $\mathbf{s}$. We denote by $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the complete-data log-likelihood function and by $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$, the expected complete-data log-likelihood function. Each iteration of the EM-algorithm involves two steps; an E-step and an M-step, defined as:

- E-step: Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ as a function of $\boldsymbol{\theta}$;
- M-step: Find $\boldsymbol{\theta}^{(r+1)}$ such that $Q(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\theta}^{(r)}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$.

Notice that, by using (15), the set-up defined above can be written as

$$\mathbf{Y}_i|t_i, u_i, \overset{\text{ind}}{\sim} N_p(\boldsymbol{\mu} + u_i^{-1/2} t_i \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, u_i^{-1}\boldsymbol{\Sigma}^{1/2}(\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)\boldsymbol{\Sigma}^{1/2}), \tag{23}$$

$$u_i \overset{\text{ind}}{\sim} h(u_i; \boldsymbol{\nu}) \tag{24}$$

$$t_i \overset{\text{iid}}{\sim} HN_1(0, 1) \quad i = 1, \ldots, n, \tag{25}$$

all independent, where $HN_1(0, 1)$ denotes the univariate standard half-normal distribution (see $|X_0|$ in equation (2) or Johnson et al., 1994). We assume that the parameter vector $\boldsymbol{\nu}$ is known. In applications the optimum value of $\boldsymbol{\nu}$ can be choosing by using the profile likelihood and the Schwarz information criterion (see Lange et al., 1989).

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ $\mathbf{u} = (u_1, \ldots, u_n)^\top$ and $\mathbf{t} = (t_1, \ldots, t_n)^\top$. Then, under the hierarchical representation (23)-(25), with $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, it follows that the complete log-likelihood function associated with $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$ is

$$
\begin{aligned}
&\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \\
&c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n} u_i(\mathbf{y}_i - \boldsymbol{\mu} - u_i^{-1/2}\boldsymbol{\Delta}t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - u_i^{-1/2}\boldsymbol{\Delta}t_i)
\end{aligned} \tag{26}
$$

where c is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. Letting $\widehat{t^2}_i = \mathrm{E}[t_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{u}_i = \mathrm{E}[U_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ $\widehat{tu}_i = \mathrm{E}[t_iU_i^{1/2}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ and using known properties of conditional expectation we obtain

$$
\widehat{t^2}_i = E[\widehat{\mu}_{T_i}^2 + \widehat{M}_{T_i}^2 + W_{\Phi_1}(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_{T_i}})\widehat{M}_{T_i}\widehat{\mu}_{T_i}], \tag{27}
$$

$$
\widehat{tu}_i = E[u_i^{1/2}\{\widehat{\mu}_{Ti} + W_{\Phi_1}(\frac{\widehat{\mu}_{T_i}}{\widehat{M}_{T_i}})\widehat{M}_{T_i}\}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i] \tag{28}
$$

where $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $\widehat{M}_{T_i}^2 = 1/(1+\widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\Delta}})$ and $\widehat{\mu}_{T_i} = u_i^{1/2}\widehat{M}_{T_i}^2\widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Gamma}}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$, $i = 1, \ldots, n$. Since the conditional expectation value defined in (27)-(28) depend only on u_i, we need to know the conditional distribution $u_i|\mathbf{y}_i$, which for this family of distributions can be easily derived from the result of the Proposition 6. Nevertheless, Monte-Carlo integration may be employed, which yield a so-called MC-EM algorithm.

It follows, after some simple algebra and using (27)-(28), that the conditional expectation of the complete log-likelihood function has the form

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}] &= c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\widehat{u}_i(\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) \\
&\quad + \sum_{i=1}^{n}\widehat{tu}_i(\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta} - \frac{1}{2}\sum_{i=1}^{n}\widehat{t^2}_i\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta},
\end{aligned}
$$

We then have the following EM-algorithm:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{t^2}_i$, $\widehat{tu}_i$ and $\widehat{u}_i$, for $i = 1, \ldots, n$ using (27)-(28).

M-step: Update $\widehat{\boldsymbol{\theta}}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ over $\boldsymbol{\theta}$, which leads to the following closed form expressions

$$
\widehat{\boldsymbol{\mu}} = \sum_{i=1}^{n}(\widehat{u}_i\mathbf{y}_i - \widehat{tu}_i\boldsymbol{\Delta})/(\sum_{i=1}^{n}\widehat{u}_i), \tag{29}
$$

$$
\widehat{\boldsymbol{\Gamma}} = \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{u}_i(\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top - 2\widehat{tu}_i(\mathbf{y}_i - \boldsymbol{\mu})\boldsymbol{\Delta}^\top + \widehat{t^2}_i\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\right],
$$

$$
\widehat{\boldsymbol{\Delta}} = \frac{\sum_{i=1}^{n}\widehat{tu}_i(\mathbf{y}_i - \boldsymbol{\mu})}{\sum_{i=1}^{n}\widehat{t^2}_i}.
$$

The skewness parameter vector and the unstructured scale matrix can be estimated by noting that $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Gamma}} + \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^T$ and $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\Delta}}/(1 - \widehat{\boldsymbol{\Delta}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Delta}})^{1/2}$. It is clear that when $\boldsymbol{\lambda} = \mathbf{0}$ (or $\boldsymbol{\Delta} = \mathbf{0}$) the M-step equations reduce to the equations obtained assuming normal/independent distribution. Note also that, this algorithm clearly generelized results found in Lachos et al. (2007, Section 2) by taken $\kappa(u) = 1$. Useful starting values required to implement this algorithm are those obtained under the normality assumption, with the starting values for the skewness parameter vector set equal to $\mathbf{0}$. However, in order to ensure that the true ML estimate is identified, we recommend running the MC-EM algorithm using a range of different starting values.

4.1 Conditional distributions for the MC-EM algorithm

In this section we compute the conditional distribution $u_i|\mathbf{y}_i$ for the distributions present in Section 3. Before, we give an important result.

Proposition 6. *(An invariance result) If $\mathbf{Y}_i \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, i=1,...,n. Then,*

$$f(u_i|\mathbf{Y}_i = \mathbf{y}_i) \propto h(u_i; \boldsymbol{\nu})\phi_p(\mathbf{y}_i|\boldsymbol{\mu}, u_i^{-1}\boldsymbol{\Sigma}). \tag{30}$$

Proof. In fact, from (19)-(21) we have that,

$$f(\mathbf{y}_i, u_i, t_i) = 2\phi_p(\mathbf{y}_i|\boldsymbol{\mu} + u_i^{-1/2}t_i\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, u_i^{-1}\boldsymbol{\Sigma}^{1/2}(\mathbf{I}_k - \boldsymbol{\delta}\boldsymbol{\delta}^\top)\boldsymbol{\Sigma}^{1/2})h(u; \boldsymbol{\nu})\phi_1(t_i|0,1)\mathbb{I}_{(t_i>0)}.$$

Hence, from Lemma 2 in Arellano-Valle et al. (2005), it follows that

$$\begin{aligned} f(\mathbf{y}_i, u_i, t_i) &= h(u_i; \boldsymbol{\nu})\phi_p(\mathbf{y}_i|\boldsymbol{\mu}, u_i^{-1}\boldsymbol{\Sigma})\phi_1(t_i|u_i^{-1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}), 1) \\ &= f(u_i)f(\mathbf{y}_i|u_i)f(t_i|u_i, \mathbf{y}_i), \end{aligned}$$

which concludes the proof. □

From Proposition 6 it follows that, under the more general SNI distribution considered here, the conditional distribution $u_i|\mathbf{y}_i$ reduces to considering the corresponding NI model. Hence, for the discussed distributions, we have the following results:

The skew-t case. If $\mathbf{Y}_i \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, $i = 1, \ldots, n$. Then, from (8) joint with the result of the Proposition (6), we have that

$$u_i|\mathbf{y}_i \sim Gamma(\frac{\nu + p}{2}, \frac{\nu + d_i}{2}),$$

so that the conditional expectation $\widehat{u_i} = \mathrm{E}[U_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ is given by $\widehat{u_i} = \dfrac{\nu + p}{\nu + d_i}$. Here, $d_i = d_{\lambda i} = (\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$.

The skew-slash case. If $\mathbf{Y}_i \sim SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, $i = 1, \ldots, n$. Then, we obtain

$$u_i|\mathbf{y}_i \sim Gamma(\nu + \frac{p}{2}, \frac{d_i}{2})\mathbf{1}_{(0,1)},$$

and $\widehat{u}_i = \left(\frac{p+2\nu}{d_i}\right)\frac{P_1(p/2+\nu+1, d_i/2)}{P_1(p/2+\nu, d_i/2)}$, where $P_x(a,b)$ denotes the cdf of the $Gamma(a,b)$ distribution evaluated on x.

The skew-contaminated normal case. If $\mathbf{Y}_i \sim SNC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu, \gamma)$, $i = 1, \ldots, n$. Then, we have that

$$h(u_i|\mathbf{y}_i) = p_i \mathbf{1}_{(u_i=\gamma)} + (1-p_i)\mathbf{1}_{(u_i=1)},$$

with

$$p_i = \frac{\nu u_i^{p/2}\exp\{-\frac{d_i u_i}{2}\}}{\nu\gamma^{p/2}\exp\{-\frac{d_i\gamma}{2}\} + (1-\nu)\exp\{-\frac{d_i}{2}\}} \text{ and } \widehat{u}_i = \frac{1-\nu+\nu\gamma^{p/2+1}\exp\{(1-\gamma)d_i/2\}}{1-\nu+\nu\gamma^{p/2}\exp\{(1-\gamma)d_i/2\}}.$$

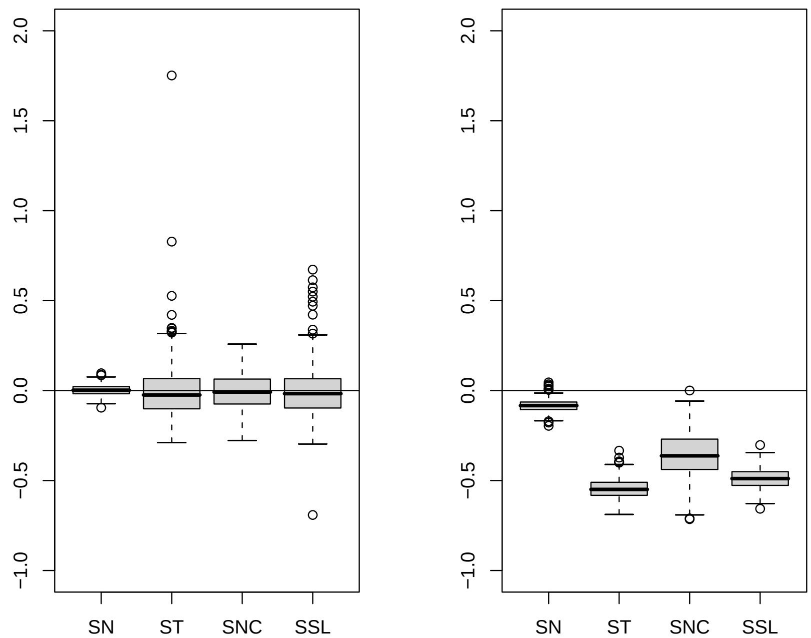
5 Applications

In this section, we present two applications. The first one illustrates the use of the distributions SN, ST, SSL and SNC in simulation studies, whereas the other one involves the statistical analysis of a real data set. The comments given here are a natural extension of those found in Wang and Genton (2006).

5.1 Simulation study

SNI distributions can be used in simulation studies as a challenging family for statistical procedure. As a illustration, we perform a small simulation to study the behavior of two location estimators, the sample mean and the sample median, under four different standard univariate settings. We consider a standard skew-normal $SN_1(3)$, a skew-t $ST_1(3,2)$, a skew-slash $SSL_1(3,1)$ and a skew-contaminated normal $SSL_1(3, 0.9, 0.1)$. The location mean of all the asymmetric distributions is adjusted to zero, so that all four distributions are comparable. Thus, this setting represents four distributions with the same mean, but with different tail behaviors and skewness. Actually, with $\nu = 1$ the skew-slash and with $\nu = 2$ the skew-t have infinite variance. We simulate 500 samples of size $n = 100$ from each of these four distributions. On each sample, we compute the sample means and the sample median and report the box-plot for each distribution in Figure 3. In the left panel, we observe that all boxplots of the estimated means are centered around zero but have larger variability for the heavy tailed distributions (skew-t and skew-slash). In the right panel, we see the boxplots of the estimated medians has a slightly larger variability that the boxplots for the estimated means at the skew-normal and skew-contaminated normal, but has a much smaller variability at the at the skew-t and skew-slash distributions. This indicates that the median is a robust estimator of location at asymmetric light tailed distributions. On the other hand, the median estimator becomes biased as soon as unexpected skewness and heavy tailed arise in the underlying distribution.

Figure 3: Boxplots of the sample mean (left panel) and sample median (right panel) on 500 samples of size n=100 from the four standardized distributions: $SN_1(3)$; $ST_1(3,2)$; $SSL_1(3,1)$, and $SNC_1(3,0.9,0.1)$. The respective means are adjusted to zero.



5.2 Fiber-glass data set

Now, the univariate skew-normal, skew-t, skew-slash and skew-contaminated normal distributions, are applied to fit the data of the strength of glass fiber, consisting of 63 observations. Jones and Faddy (2003) and Wang and Genton (2006) fit a skew-t and askew-slash, respectively, to these data. They both note skewness on the left as well as heavy tail behavior, as depicted in Figure 4. Resulting parameter estimates for the four models are given in Table 1. As suggested by Lange et al. (1989) for each model, the Schwarz information criterion (or equivalently the log-likelihood) was used for choosing among some values of ν and γ. This strategy is illustrated in Figure 5(d). Figure 4 show the histogram of the fiber data superimposed with the fitted curves of the densities for the four considered models. We have observed that the skew-contaminated normal fits the fiber data better than the other three distributions, especially at the peak part of the histogram. This conclusion is also supported by the values from the log-likelihoods given in Table 1.

Replacing the ML estimates of $\boldsymbol{\theta}$ in the Mahalanobis distance $d_i = (y_i - \mu)^2/\sigma^2$, we present Q-Q plots and envelopes in Figure 5 (lines represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points for each observation). It seems to us that the plots in Figure (5) provide even stronger evidence (than the log-likelihood criteria), that the skew-contaminated normal distribution provides a better fit to the data set than the skew-t and the skew-normal distribution.

Figure 4: The histogram of the fiber grass strength superimposed with the fitted densities curves of the four distributions.

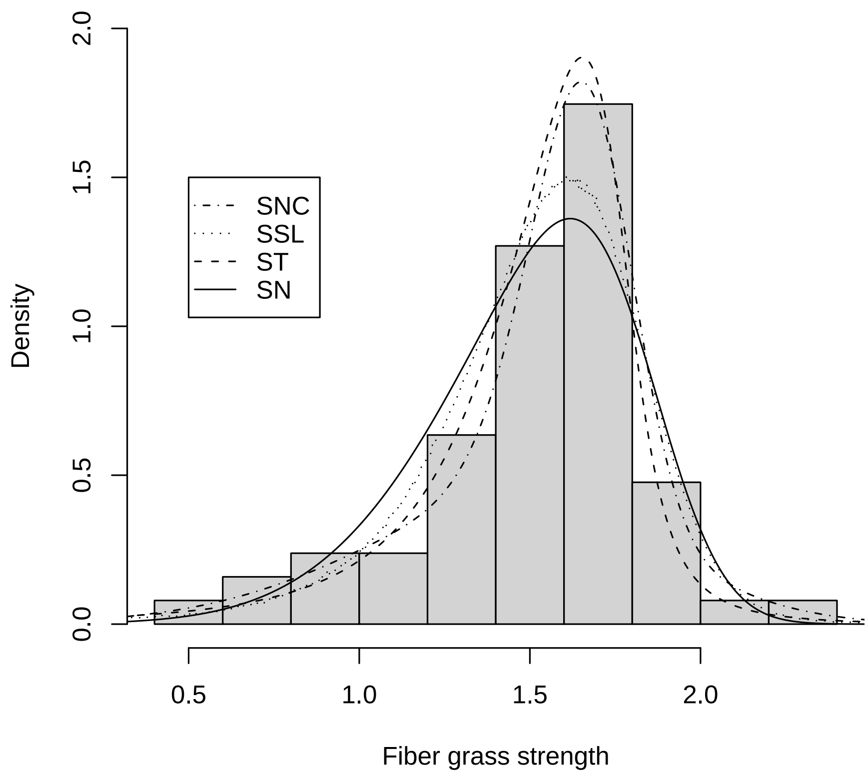


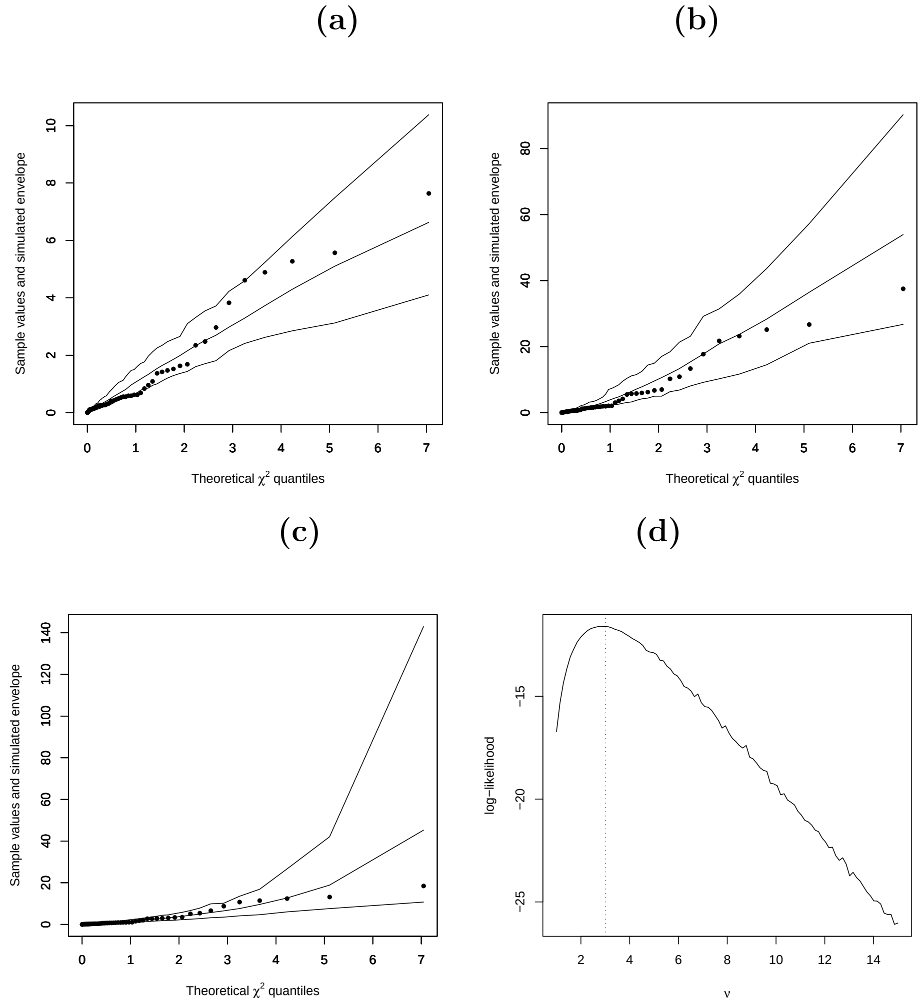
Table 1: MLE of the four models fitted on the fiber grass strength data set.

distribution	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\lambda}$	ν	γ	Log-Likelihood
SN	1.850368	0.2214105	-2.678955	-	-	-13.95719
ST	1.773221	0.08258476	-1.424273	3.0	-	-11.64703
SNC	1.763553	0.03926635	-1.304706	0.45	0.12	-10.25552
SSL	1.805591	0.08938106	-1.870298	1.73	-	-12.93673

6 Final Conclusion

In this work we have defined a new family of asymmetric models by extending the symmetric normal/independet family. Our proposal generalized recent results found in Gupta (2003) and Wang and Genton (2006). In addition, we have developed a very general method based on the MC-EM algorithm for estimating the parameters of the skew-normal/independent distributions. An important characteristic of the results obtained is that closed form expressions were derived for the iterative estima-

Figure 5: Fiber grass strength data set. Q-Q plots and simulated envelopes: (a) Skew-normal model (b) Skew-contaminated normal model (c) skew-t model and (d) profile likelihood for the skew-t model.



tion processes. This was a consequence of the fact that the proposed distributions posses a stochastic representation that can be used to represent them hierarchically. This stochastic representation also allows us to study many of its properties easily. We believe that the approaches proposed here can also be used to study other asymmetric multivariate models like those proposed by Branco and Dey (2001, Section 3). These models proposed by Branco and Dey (2001) have a stochastic representation of the form $\mathbf{Y} = \boldsymbol{\mu} + \eta(U)\mathbf{Z}$, and they also have proper elements like the skew-t, the skew-slash, the skew-contaminated normal, the skew-logistic, the skew-stable and the skew-exponential power distributions.

The assessment of influence of data and model assumption on the result of the statistical analysis is a key aspect. Work is in progress addressing specifically local influence and residual analysis.

References

Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science,* **3**, 415-438.

Arnold, B. C. and Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**, 7-54.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics,* **12**, 171-178.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, **61**, 579-602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika,* **83**, 715-726.

Branco, M. and Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis*, **79**, 93-113.

Genton, M. G., Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, **57**, 389-401.

Gupta A. K. (2003). Multivariate skew t-distributions. *Statistics,* **37(4)**, 359-363.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*, vol. 1. New York: John Wiley.

Jones, M. C. and Faddy, M. J. (2003). A skew extension of the s distribution, with applications. *Journal of the Royal Statistical Society,* Series B, **65**, 159-174.

Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. and Montenegro, L. C. (2007). Likelihood based inference for multivariate skew-normal regression models. *Communications in Statistics: Theory and Methods*, **36**. In press.

Lange, K. and Sinsheimer. (1993). Normal/independent distributions and their applications in robust regression . *Journal of Computational and Graphical Statistics,* **2**, 175-198.

Lange, K. L., Little, J. A. and Taylor, M. G. J. (1993). Robust statistical modeling using the t distribution . *Journal of the American Statistical Association,* **84**, 881-896.

Little, R. J. A. (1988). Robust Estimation of the mean and covariance matrix from data with missing values . *Applied Statistics,* **37**, 23-38.

Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with aplications to Bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129-150.

Tan, F. and Peng, H. (2006). The slash and the skew-slash Student t distributions. Submitted

Wang, J., Boyer, J. and Genton, M. (2004). A skew-symmetric representation of multivariate distributions. *Statistica Sinica* , **14**, 1259-1270 .

Wang, J. and Genton, M. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**, 209-220.

Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society,* Series B, **63**, 111-126.